

# Quelles applications mobiles poussent au renouvellement de smartphone ?

**BOUDJEMA BILAL**

**Juin 2022**

**Résumé :** Ce document présente l'analyse et la visualisation des données provenant d'une base de données comportant un grand nombre d'applications mobiles

**Mots-clés :** applications mobiles, analyse, nettoyage, visualisation, python, notebook

## 1. Introduction

Ce travail s'inscrit dans le cadre de l'unité d'enseignement Mif11 : Projet d'Orientation Master. Il s'agit d'un projet d'analyse et de visualisation de données, c'est pourquoi j'ai choisi ce thème. En fait, il y a un objectif de recherche derrière ce travail, et le sujet intéresse tous les utilisateurs de smartphones utilisant des systèmes d'exploitation comme iOS ou Android. L'objectif est de comprendre quelles applications mobiles prennent le moins en charge les appareils plus anciens.

Avec l'aide de mon encadrant, monsieur Aurélien Tabard, j'ai travaillé seul sur ce sujet.

## 2. Langage et environnement

Pour faire ce projet, monsieur Aurélien Tabard m'a demandé d'utiliser le langage Python. Ce dernier est un langage polyvalent et très flexible qui, au fil des ans, a gagné de plus en plus de crédit dans la communauté d'analyse de données.

Monsieur Aurélien Tabard m'a donné le choix d'utiliser Jupyter Notebook ou Google Colab finalement j'ai choisi ce dernier.

Google Colab [colb] fournit une plateforme pour écrire des notebooks Python contenant des éléments interactifs qui peuvent être partagés par plusieurs utilisateurs. Mais ce qui distingue Colab des autres services, c'est l'accès au GPU absolument gratuit. Cet environnement de calcul est idéal pour la visualisation des données, le calcul scientifique et le traitement des données. Des textes supplémentaires peuvent être ajoutés dans une cellule spécifique, en utilisant le format Markdown, qui introduit les problèmes et explique la solution abordée. (Figure 1).

## 3. Données :

Toutes les données proviennent de fichiers .csv (Comma-separated values) créés à partir d'une opération de web scraping qui était faite par deux autres étudiants : Popoff Hadrien et Guillardel Thomas.

Les fichiers CSV seront ensuite chargés dans des Google sheets sauvegardés et stockés dans un Google drive pour faciliter le traitement dans les prochaines opérations.

## 4. Préparation et Nettoyage de données

### 1) Préparation de données

Dans un premier temps les données venant de l'équipe de scrapping sont stockées dans des fichiers, CSV je commence à les importer dans des Google sheets et qui seront sauvegarder automatiquement à Google drive.

Après avoir inséré les données brutes dans le Google sheets l'étape de la lecture vient. À propos de cela, j'ai utilisé l'api de Google cloud plateforme en la configurant pour quelle sera adaptable avec mon compte Google afin de pouvoir faire l'authentification et la reconnaissance des Google sheets.

Voici une figure qui montre la lecture d'un fichier google sheets

### 2) Nettoyage de données :

Ultérieurement, à la suite de la lecture de données, le processus de nettoyage commence par transformer les données en data frame pour pouvoir les traiter en utilisant Pandas [pds] : une bibliothèque Python de riches de structures de données et d'outils permettant de travailler avec des ensembles de données structurés communs aux statistiques, à la finance, aux sciences sociales et à de nombreux autres domaines.

La bibliothèque fournit des routines intégrées et intuitives permettant d'effectuer des manipulations et des analyses courantes sur de tels ensembles de données. Elle a pour but d'être la couche fondamentale pour l'avenir du calcul statistique en Python.

Elle sert de complément solide à la pile Python scientifique existante, tout en mettant en œuvre et en améliorant les types d'outils de manipulation de données que l'on trouve dans d'autres langages de programmation statistique tels que R.

Ensuite, après l'exploration de la dataframe on constate qu'il y a plusieurs applications Android qui n'ont pas des valeurs (cellules vides). Ces dernières peuvent provoquer des problèmes lors de l'étape de la visualisation. De ce fait, il faudra isoler toute application qui contient une valeur manquante.

Pour cela, je les ai sauvegardées dans un fichier CSV à part afin de les envoyer à l'équipe d'extraction de données pour pouvoir récupérer les valeurs manquantes.

La colonne Version Actuelle nécessite à effectuer plusieurs traitements parmi ces derniers :

- La suppression des espaces :
- La suppression des caractères spéciaux et des lettres alphabétiques.
- Le remplacement de certains symboles par un point qui indique la version Android actuelle.

Dans le but de faciliter la visualisation et l'analyse de données. J'ai créé une nouvelle colonne sous le nom de 'Version majeure' qui sert à regrouper les différentes versions sous une seule version.

**Exemple** qui montre les versions iOS et la version majeure.

	Nom	Nom url	Categorie(s)	Date de mise a jour	Taille	Version Actuelle	Version Majeur	Version IOS	Note	Date	Heure
1	TV - Télévision Française !	968729950	News	nov. 5, 2019	20.9 Mo	2	10	10.0	1.0	2022-05-20 00:00:00	2022-06-09 10:56:37
2	ZestMoney	1597244251	Finance	May 12, 2022	59.9 Mo	1.0.7	12	12.0	1.0	2022-05-20 00:00:00	2022-06-09 11:05:00
3	Manipal Hospitals	1524836456	Medical	nov. 12, 2021	60.8 Mo	1.26	12	12.0	1.0	2022-05-20 00:00:00	2022-06-09 11:05:38
4	RTO Vehicle Registration Info	1474673684	Reference	Aug 1, 2021	25 Mo	1.4	11	11.0	1.0	2022-05-20 00:00:00	2022-06-09 11:06:38
5	Meu Imposto de Renda	922529225	Business	Apr 11, 2022	33 Mo	8.4.2000	15	15.0	1.0	2022-05-20 00:00:00	2022-06-09 11:11:21

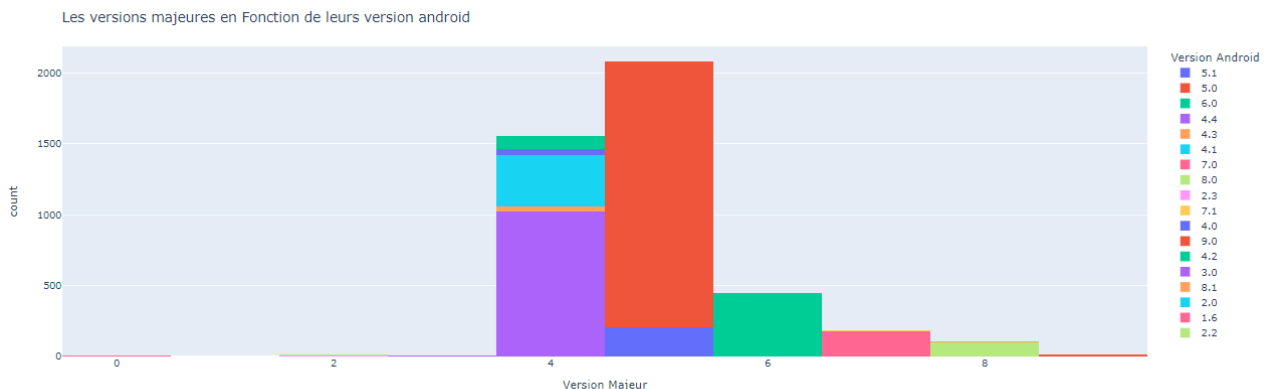
**Figure 2** : Les versions iOS et la version majeure

Finalement, la phase de nettoyage se termine par l'ordonnancement des colonnes.

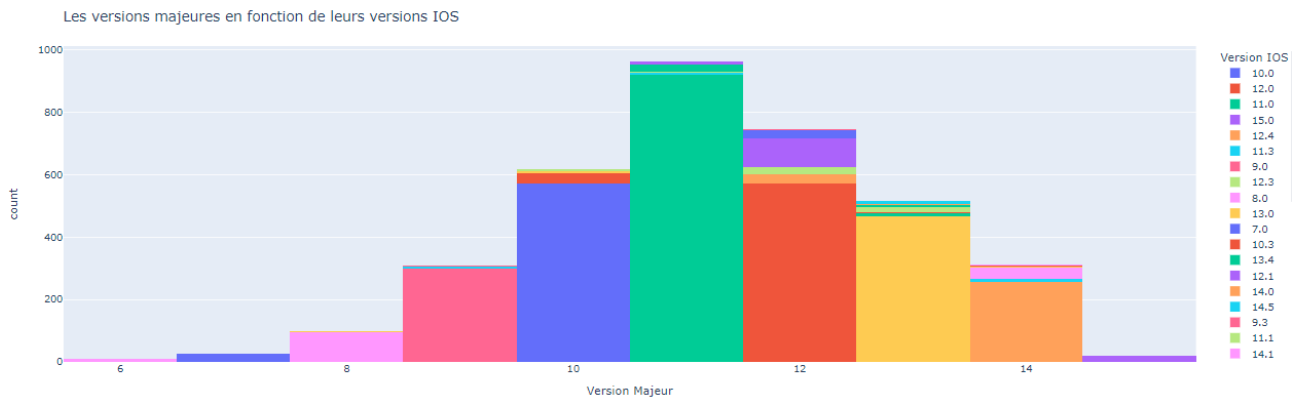
## 5. Visualisation et analyse de données

Pour la visualisation de données, monsieur Aurélien Tabard m'a donné le choix entre Matplotlib, Altair et Plotly. J'ai choisi plotly [pltly]. Plotly est une bibliothèque de visualisation interactive, cependant, elle peut devenir assez lente lorsqu'un grand nombre de points de données sont visualisés (100 000+ points de données). Cette bibliothèque résout ce problème en échantillonnage (agrégant) les données respectives à la vue et en traçant ensuite les points agrégés. Lorsque vous interagissez avec le graphique (panoramique, zoom...), des callbacks sont utilisés pour agréger les données et mettre à jour la figure.

### • Les versions majeures en fonction de leurs versions Android :



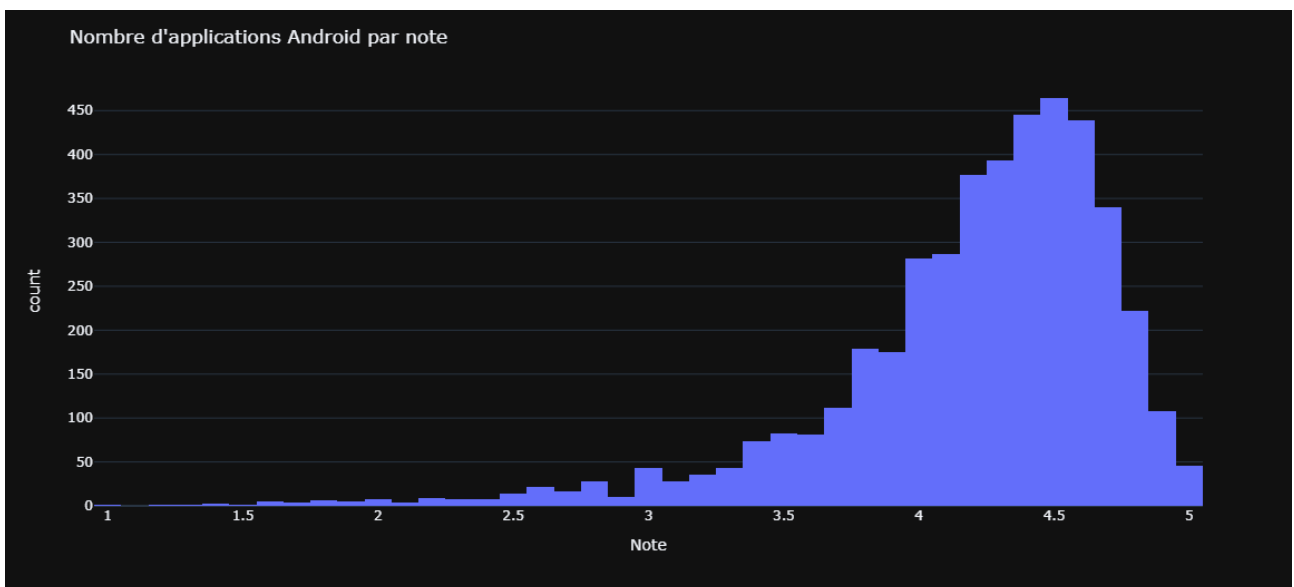
- **Les versions majeures en fonction de leurs versions iOS :**



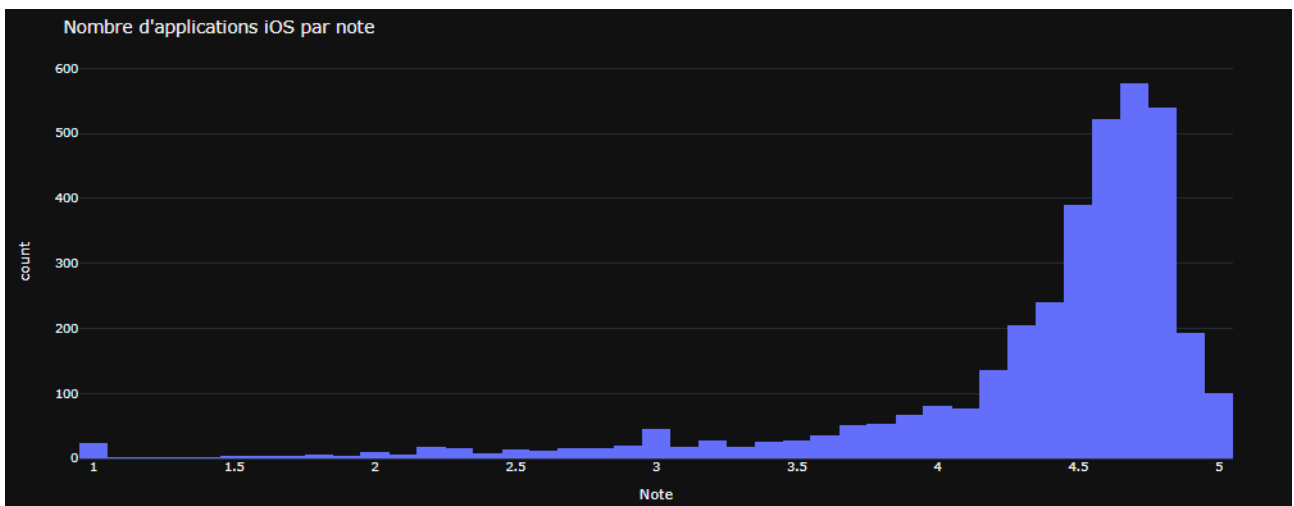
On peut extraire plusieurs informations intéressantes de ces graphiques :

Sous le système d'exploitation Android, on voit que les applications beaucoup utilisées marchent sous les anciennes versions (4,5) puis le nombre d'applications diminue au fur et à mesure des mises à jour d'Android. Par contre, les applications sous iOS on voit que les dernières versions sont beaucoup utilisées (11, 12, 13).

- **Le nombre d'applications Android par note**

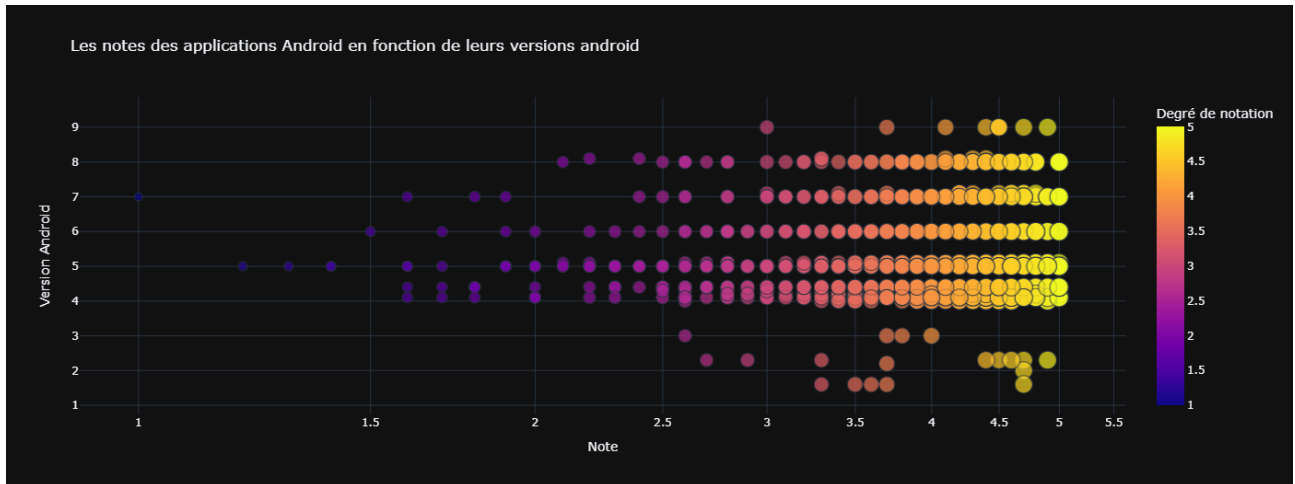


- **Le nombre d'applications iOS par note**

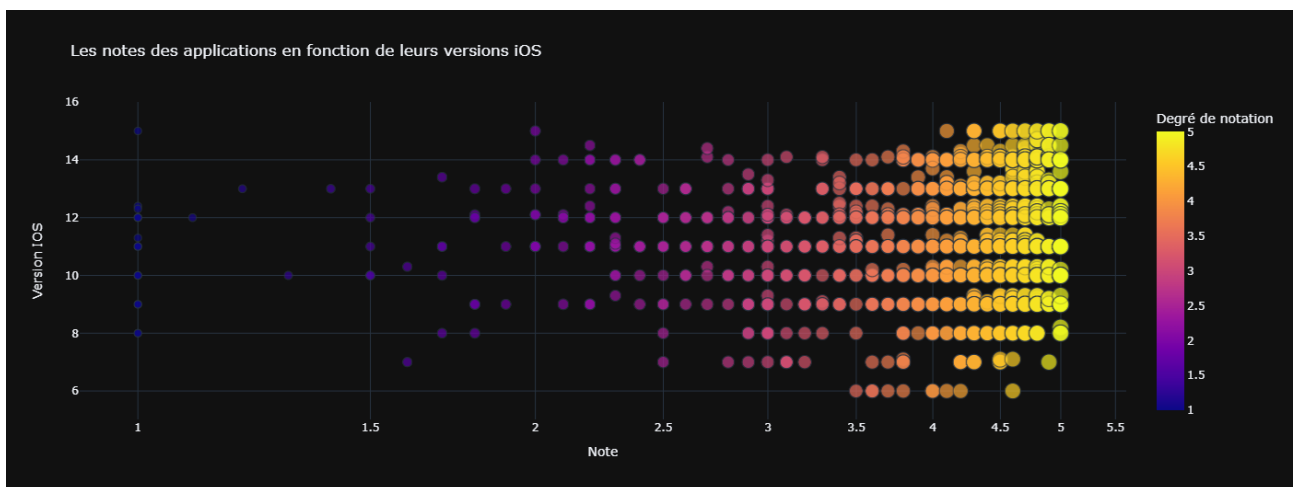


Les applications qui marchent sous iOS, on aperçoit que les applications les plus téléchargées sont bien notées (entre 4,5 et 5 étoiles) qu'Android qui possède beaucoup moins d'applications notées (entre 3.5 et 5 étoiles).

- **Les notes d'applications Android en fonction de leurs versions Android**

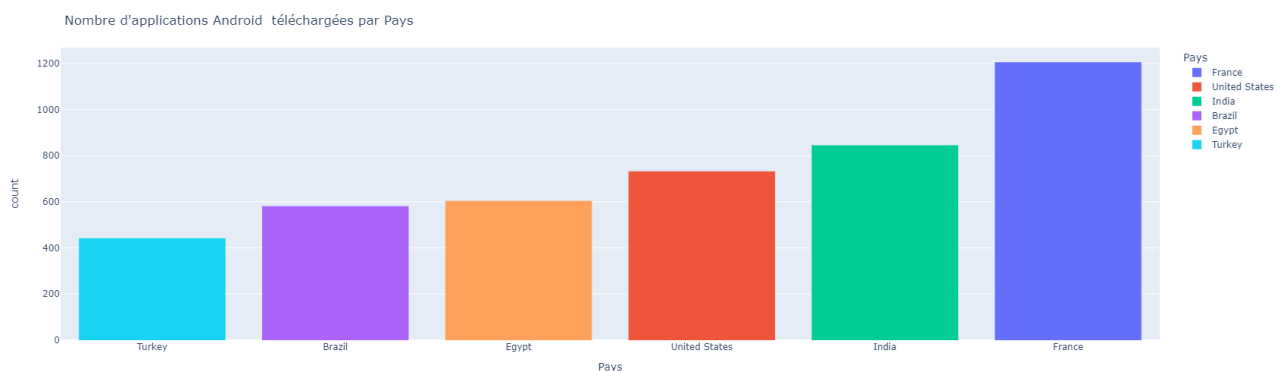


- **Les notes d'applications en fonction de leurs versions iOS**

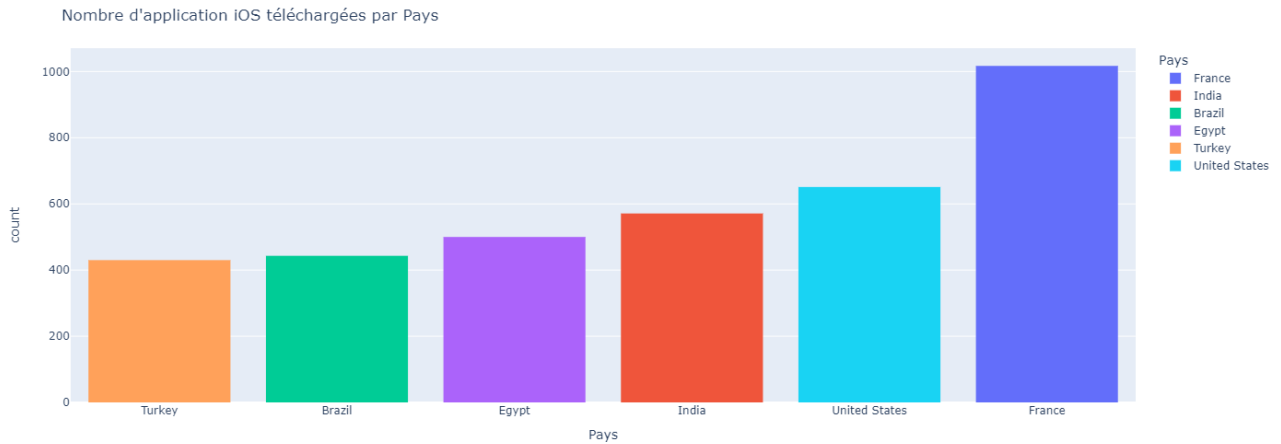


D'après ces deux graphes, on voit bien que les applications les biens notées, ce sont des applications qui marchent sous nouvelles versions par rapport aux anciennes versions.

- **Le nombre d'applications Android téléchargées par pays**

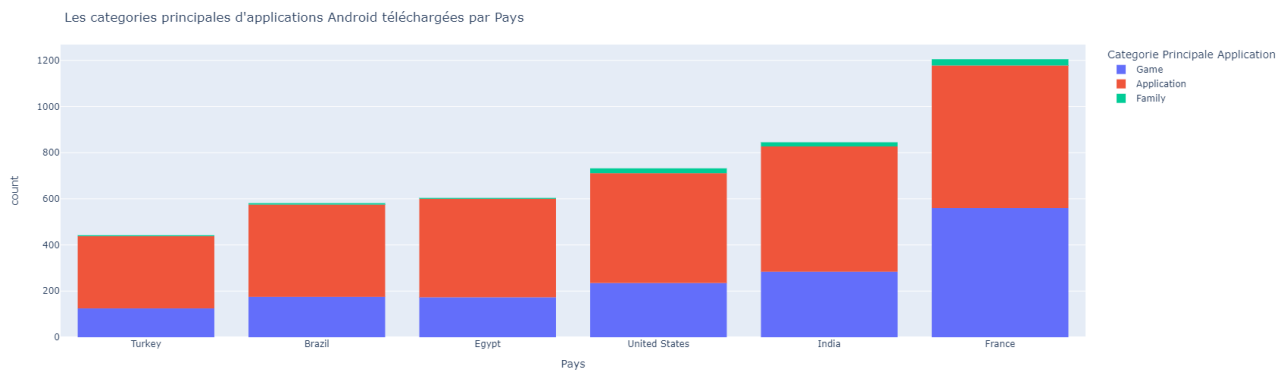


- **Le nombre d'applications iOS téléchargées par pays**



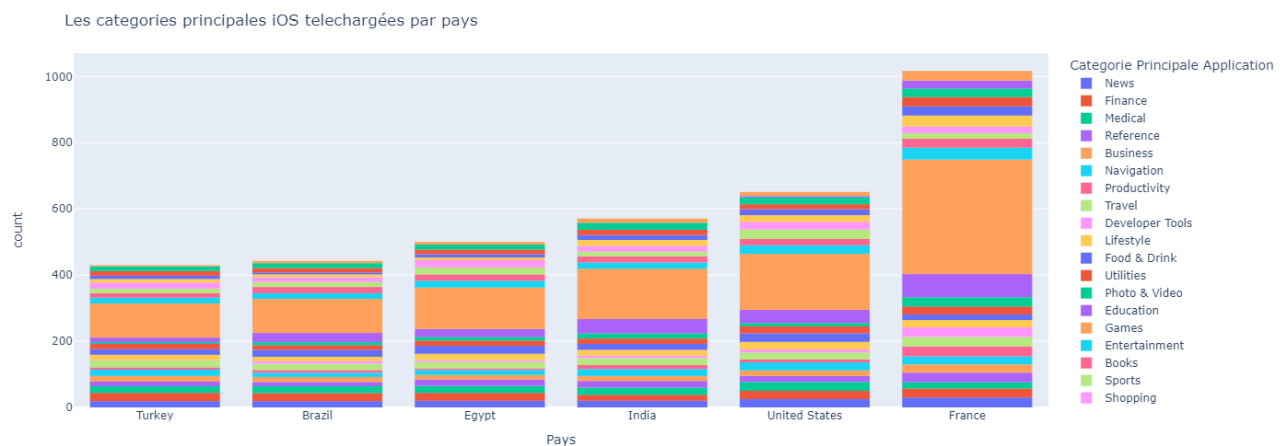
En regardant le graphe d'applications téléchargées par pays, on voit les pays développés comme la France, les États-Unis qui utilisent les applications Android même aussi les applications iOS. Ainsi que les pays qui ont une forte densité de population, comme l'Inde.

- **Les catégories principales d'applications Android téléchargées par pays**



Chez Android on remarque la catégorie application qu'elle est beaucoup téléchargées par rapport aux jeux.

- **Les catégories principales d'applications iOS téléchargées par pays**



Chez les applications iOS, les jeux la catégorie dominante au niveau de téléchargement et utilisation.

## 6. Conclusion

Dans ce projet, j'ai eu la liberté de choisir à travailler sous la plate-forme web Google Colab au lieu d'un IDE spécialisé comme PyCharm. D'autre part, travailler avec des données venant d'un binôme spécialisé en scrapping. Transmettre ces données nettoyées et organisées à l'équipe de développement du site Web.

Dans l'ensemble, ce projet m'a permis de solidifier certains acquis, notamment au niveau du langage de programmation python. Utiliser des bibliothèques comme pandas et plotly, mais aussi apprendre à être rapide et prêt face à des données qui peuvent changer leurs structure et quantité à tout moment.

## Références

[colb] : Google collaboration plateforme : <https://colab.research.google.com>

[pds] : bibliothèque d'analyse complexe de données : <https://pandas.pydata.org>

[pltly] : bibliothèque de visualisation interactive : <https://plotly.com/python/>